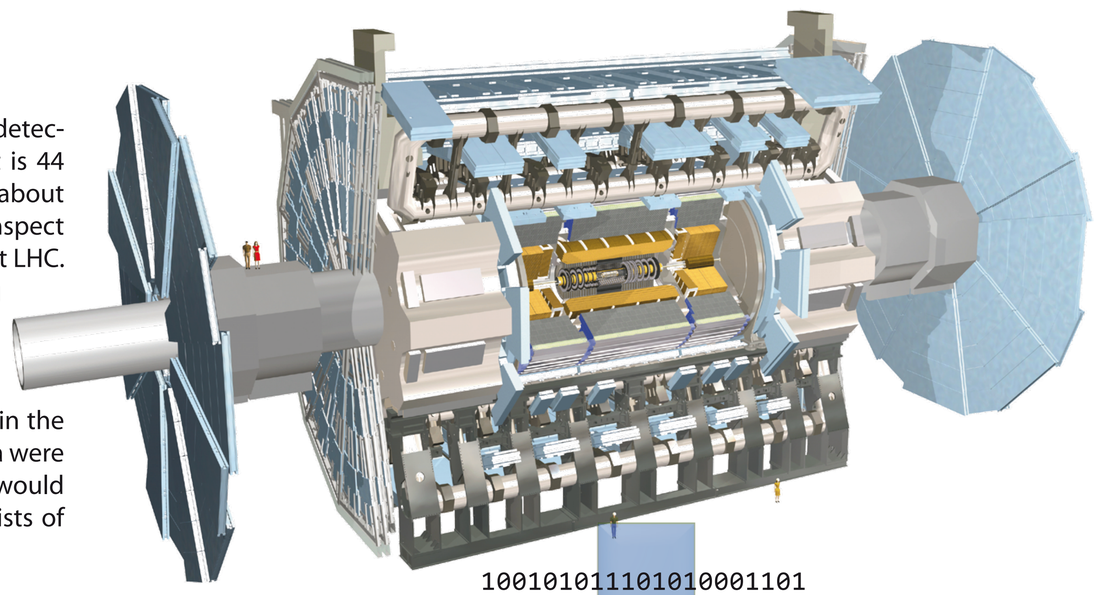
DATA PROCESSING IN ATLAS

Tomáš Kubeš
Institute of Physics of the AS CR
on behalf of the ATLAS Collaboration

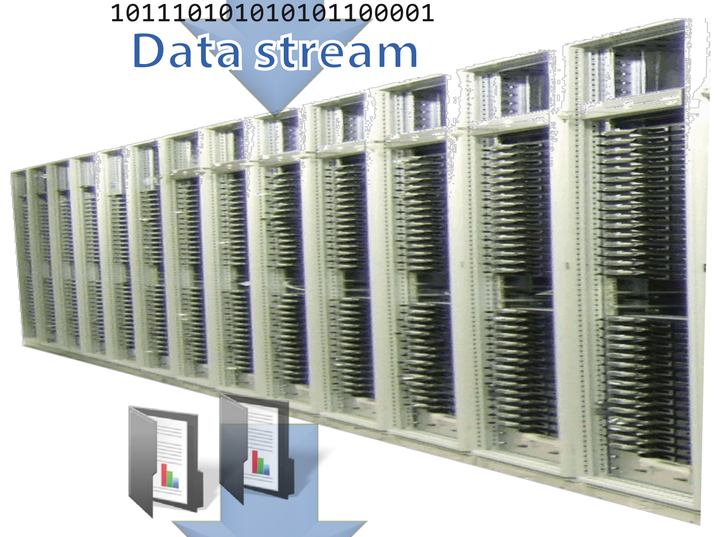
ATLAS Detector

ATLAS (A Toroidal LHC ApparatuS) is one of the four detectors at the Large Hadron Collider (LHC) in CERN. It is 44 metres long and 25 metres in diameter, weighing about 7,000 tonnes. It is a general purpose detector to inspect particle collisions (mainly proton-proton) occuring at LHC. The detector consists of an inner tracker measuring the momentum charged particle, calorimeter for measuring energies carried by particles, and muon spectrometer for measuring properties of muons. About 40 million beam crossings per second occur in the center of the detector at full luminosity. If all the data were to be saved, ~100 petabytes per second of raw data would have to be stored. Each event in the detector consists of about 25 MB of raw data (compressible to 1,5-2 MB).



10010101110101010001101
10111010101010100001

Data stream



RAW data files

Throughput from EF to T0: 320 MB/s

Level 1 Trigger

The trigger system selects events of physics interest for further processing. The first level trigger is massively parallel system based in the special electronics of the detector itself. It works in the scope of subdetectors. Decision speed is less than 2,5 μs. About 100 000 out of 40 000 000 bunch collisions per second pass this trigger level and are sent further.

Level 2 Trigger

2nd level trigger runs on a cluster of 500 machines which select events on the base of detailed analysis of the regions of interest passed by Level 1 Trigger. The Level 2 Trigger works with combined region of interest information and full granularity data from all subdetectors. Processing time is 40 ms. Less than 2 000 events pass the 2nd level trigger.

Event Filter

Event filter is an ATLAS level 3 trigger. It runs on the Event Filter Farm where detailed analysis of all built events passed by the level 2 trigger is performed (less complex and faster algorithms than in off-line analysis are used, but condition data are avialable). Decision speed is on the order of 1-4 s. Less then 200 events per second are selected for storing and further offline analysis and sent to Tier-0 by Sub-Farm Output manager for further data processing.

Sub-Farm Output Manager

Set of 5 custom designed file servers with special filesystem which collect data passed by the Event Filter and store them in form of data files to a dedicated CASTOR Tier-0 pool. It also saves information to the handshake table (realized through Oracle db.) to notify Tier0 that data are ready to be processed.

Event Filter Farm

Event Filter Farm is a large cluster of computers located near the experiment. It takes care of reading the detector outputs, event assembly, and last stage event filtering (3rd level trigger). For all events selected by the 2nd level trigger, event builder processor collects all fragments from input buffers into a single memory and assembles them into files. Then scaled down versions of off-line analysis algorithms are applied, they can access basic calibration and alignment ("condition") data. The processing time per event could be about 1-4 seconds on a 1000 MIPS computer. Sub-Farm Output Manager then collects files for selected events and ships them to a CASTOR pool for further processing by Tier-0 and beyond.

CASTOR

CASTOR, stands for the CERN Advanced STORage manager. CASTOR allows storing of huge amounts of data transparently by seamlessly managing disk cache(s) and tape storage. ATLAS has several PB (petabytes) of disc cache storage in various pools and "unlimited" amount of tape storage. Castor pools are used to store both final and intermediate products during the ATLAS data processing.

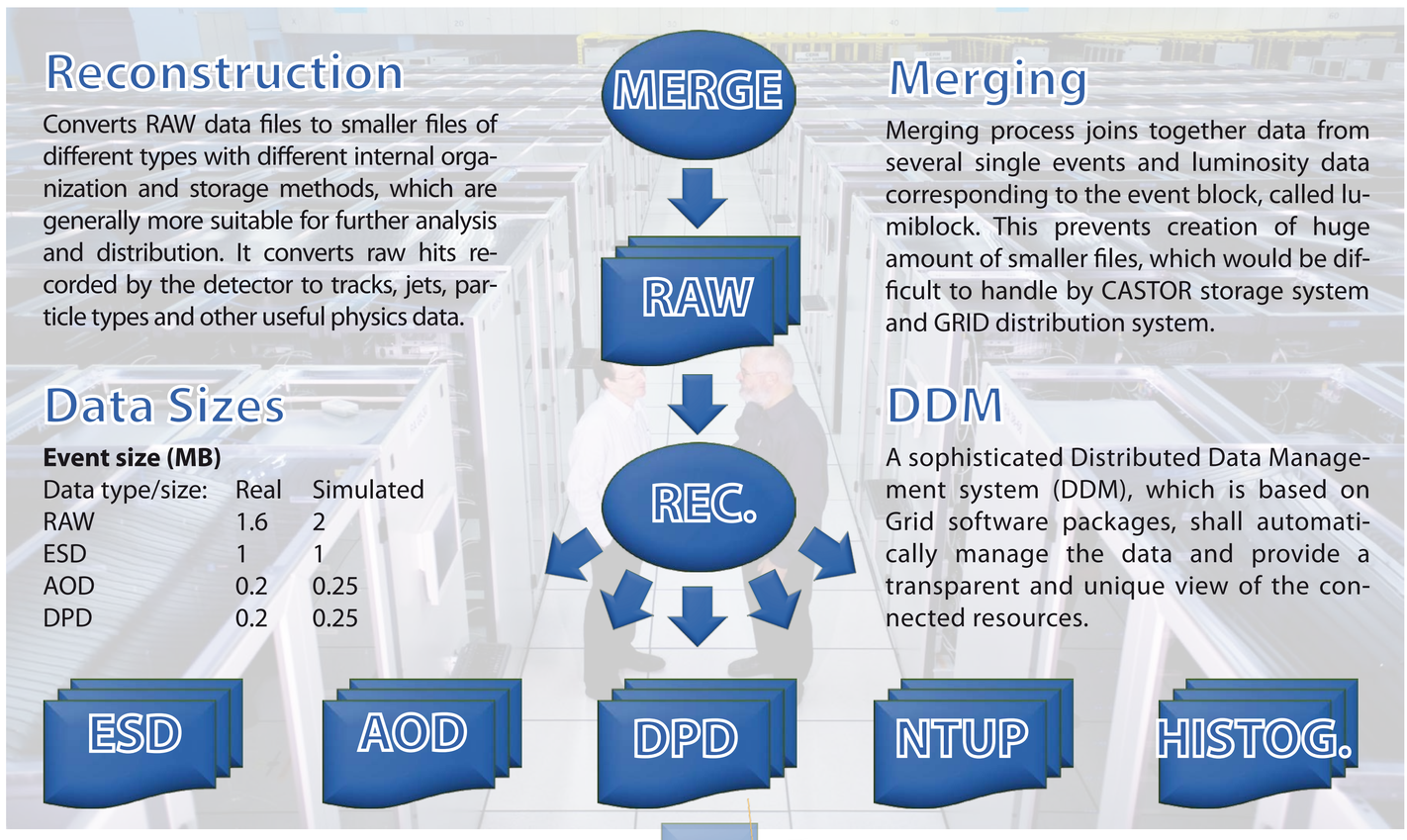
ATLAS Tier-0

The task of the ATLAS Tier-0 system is to perform the prompt first pass processing on the express/calibration physics stream, 24-48 hours processing of full physics data stream with reasonable calibrations, and to register raw and reconstructed data to the Distributed Data Management System (DDM) which then distributes them to the Tier-1 centres and beyond. ATLAS Tier-0 is composed of roughly 100 interconnected powerful computers housed in the Cern Computing Center.

GRID

The LHC Computing Grid is a distribution network designed by CERN to handle the massive amounts of data produced by the LHC. It incorporates both private fiber optic cable links and existing high-speed portions of the public Internet. It is hierarchically structured to Tier-0, Tier-1 and Tier-2 centers. It allows users to send their analysis jobs to places where data reside seamlessly.

ATL-SOFT-SLIDE-2009-329
21 October 2009



How Tier-0 works?

Once new RAW data arrive to Tier-0 pool, information from the handshake table is picked up by TOM (Tier-0 management system), which acts as an orchestrator on Tier-0. Based on the data arriving, TOM will define and follow up upon all tasks necessary to perform the prompt reconstruction. TOM will also upload all resulting datasets into the ATLAS distributed data management system. The actual execution of the defined tasks is then handled by a dedicated instance of the supervisor executor (Eowyn). T-Zex (Tier Zero Executor) then actually submits computing jobs corresponding to each task to the Tier-0 local LSF farm and informs Eowyn once they are completed or failed. Monitoring system is build on the top of the system to give a full image of the state of Tier-0, tasks, jobs, datasets, and necesssary critical resources like CASTOR pools and upload ques.

User Data Analysis on the GRID

Physics analysis can be carried out in the ATHENA framework and GRID submissions can be handled through GANGA, both are developed at CERN.

Data Types

Raw Data (RAW) contain all information from the detector in compressed form. Each run and each stream produce one dataset (logically connected files).

The **Event Summary Data (ESD),** are processed RAW data which still contain sufficient information to re-run parts of the reconstruction, such track refitting, jet calibration, etc. They should be used preferentially to RAW data.

The **Analysis Object Data (AOD),** which consists of a reduced size output of physics quantities from the reconstruction that should suffice for most kinds of analysis work. Separate tailor-made streams of AODs are foreseen for many different needs of the physics community.

Derived Physics Data (DPD) can be either subsets of ESDs that can be used to study detector performance: "Performance DPDs", or "Physics DPDs", which are derived from AODs and are useful for specific analysis.

Ntuples and **Histograms** are special kind of accompanying data.

Datasets with events

Tier-1

Tier-1 centers are very large computing centers for a whole country or region, each is connected to CERN by 10 GBit/s line.

Tier-2

Tier-2s are medium local computing centers at institutes, they serve to end users and support additional small Tier-3 centers.

Data Distribution Policy

RAW (Raw Data)
1 copy of RAW data is kept at CERN on tapes.
1 copy of RAW data is distributed among Tier-1 centers and moved to tapes.
Samples can be moved to Tier-2s for detailed analysis.
ESD (Event Summary Data)
1 copy is kept at CERN for the first-pass processing and then moved to tapes.
1 full copy is moved to Brookhaven National Laboratory.
2 copies are distributed to random Tier-1 centers (ESD always follow RAW data), ESDs are kept on discs. Parts can be copied to Tier-2s for analysis.
AOD (Analysis Object Data)
1 copy is stored at CERN on tapes.
1 copy goes to every Tier-1 cloud where it is stored on discs, exact distribution between Tier-1 and Tier-2s in the cloud can vary.
DPD (Derived Physics Data)
Distributed the same way as AODs.

SYMMETRIES AND SPIN (SPIN-Praha-2009)
Prague, July 26 - August 2, 2009